



AVVISO CONSULTAZIONE PRELIMINARE DI MERCATO

Avviso di consultazione preliminare di mercato propedeutica all'indizione di una procedura negoziata senza previa pubblicazione di bando di gara per l'acquisizione di sistemi di intelligenza artificiale e analisi semantica per la lingua italiana e per la lingua inglese dotati di tecnologia linguistica semantica contenente analisi morfologica, tagging grammaticale, analisi logica e disambiguazione dei termini, per l'elaborazione su basi di dati anche non strutturati, per l'estrazione dei significati e la comprensione del linguaggio naturale.

S'informa che l'Agenzia per l'Italia Digitale (AgID) intende avviare una procedura negoziata senza previa pubblicazione di bando di gara per l'acquisizione di sistemi di intelligenza artificiale e analisi semantica.

Il presente Avviso persegue le finalità di cui all'art. 66, comma 1, del decreto legislativo n. 50/2016 (Codice degli appalti) ed è volto – sulla base delle indicazioni fornite dall'Autorità nazionale anticorruzione (ANAC) – a confermare l'esistenza dei presupposti che consentono, ai sensi dell'art. 63 del Codice degli appalti, il ricorso alla procedura negoziata in oggetto, ovvero ad individuare l'esistenza di soluzioni alternative per l'acquisizione di sistemi di intelligenza artificiale e di analisi semantica, dotati delle caratteristiche e degli strumenti individuati dall'AgID e specificati nella scheda tecnica che segue.

Gli operatori del mercato che ritengano di poter fornire il sistema di intelligenza artificiale e di analisi semantica rispondente al fabbisogno e agli strumenti individuati dall'AgID, ovvero di suggerire e dimostrare la praticabilità di soluzioni alternative, dovranno far pervenire la propria offerta in relazione alla scheda tecnica che segue **entro e non oltre le ore 12.00 del giorno 10/04/2017** all'indirizzo PEC, protocollo@pec.agid.gov.it.

Tale PEC dovrà riportare in oggetto la seguente dicitura: <<**Risposta a consultazione preliminare di mercato propedeutica all'indizione di una procedura negoziata senza previa pubblicazione di bando di gara per l'acquisizione di sistemi di intelligenza artificiale e analisi semantica**>>.

La partecipazione a detta consultazione non determina aspettative, nè diritto alcuno e non rappresenta invito a proporre offerta, nè impegna a nessun titolo l'AgID nei confronti degli operatori interessati, restando altresì fermo che l'acquisizione del sistema d'intelligenza artificiale e analisi semantica oggetto della presente consultazione è subordinata all'apposita procedura che sarà espletata dall'AgID ai sensi del richiamato decreto legislativo n. 50/2016.

*** **

Scheda Tecnica – Requisiti del sistema di intelligenza artificiale e di analisi semantica per la lingua italiana e per la lingua inglese dotati di tecnologia linguistica semantica contenente analisi morfologica, tagging grammaticale, analisi logica e disambiguazione dei termini, per l'elaborazione su basi di dati anche non strutturati, per l'estrazione dei significati e la comprensione del linguaggio naturale.

1. Fabbisogno

Il CERT-PA ha l'esigenza di integrare la propria struttura di analisi delle minacce cibernetiche attraverso un sistema di intelligenza artificiale e analisi semantica al fine di:

1. monitorare le informazioni non strutturate, provenienti da OSINT (siti web, blog, feed news, social network, ecc..) e CLOSINT (canali come IRC e Pastebin);
2. classificare automaticamente le informazioni per l'estrazione di entità e relazioni a partire dalle informazioni analizzate;
3. predisporre delle tassonomie personalizzate per le specifiche esigenze del CERT-PA;
4. fornire supporto al lavoro dell'analista nella gestione dell'intero processo di analisi delle informazioni (definizione fonti e tassonomie di riferimento, acquisizione, selezione, gestione alert, valutazione e validazione, disseminazione);
5. disporre di un'interfaccia web interattiva, comprensiva di visualizzazioni grafiche a supporto del lavoro di analisi.

Tale iniziativa è necessaria per permettere al CERT-PA di svolgere i compiti istituzionali ad esso assegnati, in particolare per quel che riguarda l'erogazione alle pubbliche amministrazioni dei servizi proattivi, atti ad anticipare per quanto possibile l'adozione degli strumenti idonei a contrastare e mitigare gli attacchi.

2. Durata e pianificazione delle attività

Durata delle attività: da maggio 2017 a dicembre 2019.

3. Strumenti, soluzioni tecniche individuate, requisiti

Sistema di intelligenza artificiale e analisi semantica per la lingua italiana e la lingua inglese personalizzabile dall'utente finale e caratterizzata da una analisi linguistica ascrivibile come 'deep semantic'.

In particolare, detto sistema deve essere dotato di tecnologia linguistica semantica contenente analisi morfologica, tagging grammaticale, analisi logica e disambiguazione dei termini, per l'elaborazione su basi di dati anche non strutturati, per l'estrazione dei significati e la comprensione del linguaggio naturale.

Il sistema deve disporre dei seguenti requisiti e funzionalità:

- utilizzo di una rete semantica estremamente ricca e organizzata in termini di concetti e di relazioni concettuali;
- analisi morfologica e tagging grammaticale;
- utilizzo di tecnologia linguistica semantica contenente l'analisi logica e la disambiguazione dei termini per l'elaborazione su basi di dati anche non strutturati, per l'estrazione dei significati e la comprensione del linguaggio naturale;
- funzione di text mining, con capacità di estrarre sia entità standard (persone, aziende, enti/organizzazioni, località) sia entità custom (concetti dell'ontologia di dominio caratterizzante i contenuti oggetto di analisi, come ad esempio il mondo cyber), con estrazione automatica anche di relazioni tra le suddette entità ed integrazione con regole di estrazione ad hoc realizzabili direttamente dal CERT-PA, con l'uso di specifici tool per la personalizzazione della rete semantica e delle modalità di funzionamento della categorizzazione e del text mining (estrazione).
- moduli di analisi semantica per la classificazione automatica di contenuti testuali multilingue;
- un potente motore di analisi in grado di evidenziare automaticamente alert e warning;
- un insieme di tool per la visualizzazione grafica dei risultati;
- logiche di proposizione dei risultati con ranking personalizzabile;
- analisi interattiva e correlazione dei dati, strutturazione automatica di entità e relazioni;
- strumenti per la personalizzazione della componente di conoscenza semantica e per la definizione di regole di text mining e di classificazione;
- possibilità di acquisire i dati da tutti i principali database e di generare analisi di scenario;

- sistema di traduzione automatica.

La soluzione individuata dovrà inoltre essere integrata con le tassonomie individuate come segue:

- **Tassonomia Intelligence**
Dovrà essere strutturata su almeno 15 categorie di primo livello, articolate in sotto-categorie di dettaglio organizzate su 3 livelli, la tassonomia Intelligence costituisce una verticalizzazione della tassonomia IPTC (www.iptc.org) specifica per il dominio intelligence. L'albero tassonomico dell'IPTC deve essere implementato con categorie afferenti a domini di settore quali intelligence, difesa, sicurezza interna, sicurezza internazionale, infrastrutture critiche. Il motore di categorizzazione basato sulla tassonomia Intelligence deve coprire almeno 500 categorie.
- **Tassonomia European Crime**
Tassonomia relativa al dominio *crime* basata sullo standard europeo definito dalla "Swedish Initiative", iniziativa nell'ambito della quale sono stati condivisi un protocollo e un modello di conoscenza che favorissero lo scambio di informazioni e di dati strategici tra le autorità e le forze dell'ordine degli Stati europei. La tassonomia consta di 32 categorie, organizzate su un unico livello, ciascuna dedicata ad un crimine o ad una famiglia di reati.
- **Tassonomia Cyber Crime**
La tassonomia dedicata al dominio "cyber crime" deve disporre di almeno 20 categorie, articolate su tre livelli. La tassonomia deve focalizzarsi su due macro-temi principali, l'illecito informatico e la sicurezza informatica. Per quanto riguarda gli illeciti informatici, le categorie previste sono, ad esempio, attacchi informatici (dettagliato in Attacchi DoS, Intrusione informatica), frodi informatiche, hackers; per quanto riguarda la sicurezza informatica, i temi trattati riguardano le vulnerabilità, i malware e i dispositivi di protezione.

Il CERT-PA dovrà avere la possibilità di gestire in piena autonomia le tassonomie e le altre personalizzazioni linguistiche implementate dal sistema. In particolare dovrà essere predisposto un ambiente di sviluppo desktop delle personalizzazioni linguistiche per creare nuove tassonomie ed estrazioni di entità.

Dovrà essere possibile:

- definire un elenco di campi e valori (copie Tipo/Valore) di interesse;
- scrivere regole di estrazione e categorizzazione;
- consultare e interrogare la rete semantica;

- verificare la validità sintattica delle regole;
- eseguire test di precision e recall.

4. Costi attesi

Costi software piattaforma analisi semantica, non superiori a 320.000 euro; costi licenze software annuali per sistemi di traduzione automatica, non superiori a 30.000 euro; costi concernenti servizi professionali per attività di personalizzazione e sviluppo, non superiori a 270.000 euro

*** **

Il Responsabile del procedimento è l'ing. Mario Terranova, Dirigente Responsabile dell'Area Sistemi, tecnologie e sicurezza informatica dell'AgID: e-mail: terranova@agid.gov.it .

Roma, 24 marzo 2017

Antonio Samaritani